AD_____

Award Number: DAMD17-98-1-8039

TITLE: FACTS (Find the Appropriate Clinical Trials) for You: A Computer-Based Decision Support System for Breast Cancer Patients

PRINCIPAL INVESTIGATOR: Lucila Ohno-Machado, M.D., Ph.D.

CONTRACTING ORGANIZATION: Brigham and Women's Hospital
Boston, Massachusetts 02115

REPORT DATE: May 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 074-0188*



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | May 2000 | Annual (20 Apr 99 - 19 Apr 00) |

**4. TITLE AND SUBTITLE**
FACTS (Find the Appropriate Clinical Trials) for You: A Computer-Based Decision Support System for Breast Cancer Patients

**5. FUNDING NUMBERS**
DAMD17-98-1-8039

**6. AUTHOR(S)**
Lucila Ohno-Machado, M.D., Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Brigham and Women's Hospital
Boston, Massachusetts 02115

**E-MAIL:**
machado@dsg.harvard.edu

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** ***(Maximum 200 Words)***

We report on accomplishments of Year 2 of a three-year award.
In year 1, we have built a deterministic computer-based system to help patients and physicians select appropriate clinical trials for patients with advanced breast cancer.
In year 2, we refined the deterministic system and introduced management of uncertainty in the form of belief networks. These networks are used to estimate values for unknown eligibility criteria given values for related variables. They are implemented in C++ and utilize belief-network software commercially available. We have utilized accepted standards for representing medical terminology and for expressing eligibility criteria in a computer-readable format. In particular, we have extended Arden syntax to deal with particular requirements of this application.
In year 2, we redesigned the evaluation plans and propose a clinical trial for evaluation of the system in year 3. In this trial, a comparison of oncologists' performances with and without the system will be performed. The detailed plan, as well as its rationale is attached.



| 14. SUBJECT TERMS | 15. NUMBER OF PAGES |
|---|---|
| Breast Cancer | 29 |
| | 16. PRICE CODE |

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

# FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

N/A Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

N/A Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

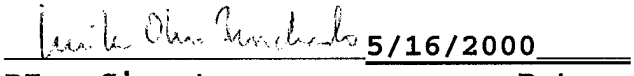
N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.



5/16/2000

PI - Signature Date

# Table of Contents

# Introduction

This report refers to the second term of a three year award. The main accomplishments of this year, relative to the last progress reports are outlined in the body of this document.

Although participation in clinical trials has been shown to improve health outcomes, accrual of patients is difficult and is estimated to be below 5% of the eligible population. Lack of information and automated tools to search clinical trials appropriate for each particular patient are some of the main reasons for low accrual. The purpose of this project is to build and evaluate a computer-based decision support system to help patients and primary care providers seek appropriate trials for their specific situation, even in conditions of uncertainty (missing data). We have proposed to make available, via the WWW, a search engine for clinical trial eligibility that searchers trials listed in the PDQ database of the NCI. On-line description of the project and working prototype can be found from http://dsg.harvard.edu/public/dsg/projects/facts.html

# Body

Briefly, we have proposed to build our computer-based eligibility determination engine in two stages: (1) build an ad-hoc deterministic (i.e., nonprobabilistic engine not able to deal with uncertainty or consider associations among eligibility criteria and patient data values), and (2) build a probabilistic engine, based on belief networks, that is able to statistically imput values for missing data, given the information it can gather from the patient or health care provider, and can take into account associations among variables and patient data values.

An overall summary of the goals and accomplishments for the first year of this project is given in [Ohno-Machado et al., 1999]. Additional work has been developed since then, and is included in the summary below, in which a description of the research accomplishments associated with each Task outlined in the Approved Statement of Work (restated in **bold** face) are given:

***Task 1.*** **Analyze, structure, and construct data entry forms for eligibility criteria derived from clinical trials for breast cancer treatment available in PDQ, Months 1-6 (ACCOMPLISHED):**

a. **PDQ clinical trial summaries for health care professionals will be dissected**

In the first year of this proposal, we have analyzed and encoded 85 clinical trials from the PDQ database. These are all Phase II and Phase III trials for the treatment of metastatic or recurrent breast cancer. A total of 2188 criteria in these trials were encoded. The medium number of crtieria per protocl was 25, with numbers ranging from 6 to 45. The encoded criteria were structured in the format described in the next item. Approximately half of the protocols available in year 1 have changed their status for year 2 (i.e., as many as half

of the protocols have been closed to new participants. In addition, new protocols have been added. We are currently encoding the new protocols and updating the status of the old ones. We did not anticipate such a dynamic environment and have therefore extended task 1 to last for the 30 first months of this project. Although this is not necessary to make the "proof of concept" proposed in this research project, it is desirable in order to make the system useful in the real world.

Examples of encoded protocols and of variables utilized in coding all protocols were given as appendices in the last progress report.

**b. A structured format for storing eligibility criteria in a relational database will be defined**

In the first year of this project, we have evolved three different structure formats for storing eligibility criteria into databases. In the second year, we opted towards using exclusively an extension of Arden syntax, currently supported by the InterMed collaboratory and its GuideLine Interchange Format (GLIF) specification [Peleg et al, 2000, in *Appendix 1*]. The extension to Arden, briefly documented in [Wang et al, 1999], was created to circumvent these problems. Since the last progress report, we have incorporated all operators from Arden syntax. We used the eXtensible Markup Language (XML) to represent the structure of the protocols in terms of objects and attributes. Examples of the Document Type Definition for the protocol files and the variables utilized in them are were reported last year. A relational database was not necessary to store the eligibility criteria, as the XML files were deemed more general and could be parsed in real time with no performance degradation.

**c. WWW-based data entry forms will be constructed an linked to database**

In year 1, we developed static forms that were connected to a database and used the WWW interface to acquire information from patients. In year 2, we made the forms more dynamic and linked to XML files and our deterministic and probabilistic engines.

**d. Database for interim storage of patient data will be constructed**

A simple structured format was used to store patient data and communicate it to the WWW-server. We used XML files for this purpose. Since we plan to link to a real patient database in the future, and this database is supposed to be interim storage, lasting one WWW session only, detailed elaboration of structure for the patient data was not necessary.

***Task 2*. Construct simple models that do not model uncertainty to assess the need for belief network models, Months 7-9 (ACCOMPLISHED):**

a. **Simple rule-based system construction using knowledge from domain expert**

In year 1, we constructed a deterministic system. The rule-based system used for this phase consisted of the following. Protocols for which the patient had one or more values that met exclusion criteria (or did not meet inclusion criteria) were removed from consideration for that patient. For the remaining protocols, the ranking took into account how many criteria still needed to be met. A protocol with several criteria to be met would rank lower than one with just one or two criteria. The importance of each criterion was not taken into account, nor the dependencies among criteria.

**b. Preliminary evaluation of simple rule-based system**

We have consulted with our two oncologists to assess how this simple system was performing, both in terms of adequacy of trials selected and usability of the interface. The experts provided useful suggestions, which were incorporated. We have also presented the system to a few colleagues, obtaining positive feed-back. We have designed a clinical trial to allow better evaluation of the deterministic system [Lacson et al, 2000 in *Appendix 2*]. We have collected patient data from medical records to implement this trial.

***Task 3.*** **If results from Task 2 show that belief networks are needed, construct belief network to model uncertainty in most common eligibility criteria and perform inference on entered data, else refinement of simple models and interface construction will take place, Months 9-12 (ACCOMPLISHED):**

**a. Belief network model will be constructed using knowledge from domain expert**

Simple Belief networks featuring relations among laboratory values that are frequently encountered in eligibility criteria were constructed by Dr. Huan Le, an internist and current postdoctoral fellow in medical informatics. The small belief networks deal with few demographic data and laboratory values related to liver, renal, and hematologic function.

**b. Belief network model will be integrated with WWW and database environments to create application**

We have incorporated the Belief network software acquired for this project, Netica, from Norsys Inc., into the current eligibility engine. This incorporation required extensive software engineering, given the complexity of the Arden implementation and the requirements of the ranking algorithm.

**c. Algorithm for ranking possible trials for a patient will be implemented**

The ranking algorithm was based on the pre-existing one, in which protocols are ranked in terms of how much more information needs to be collected for a certain patient to be deemed "appropriate". This algorithm, however, incorporated the imputation of missing values as provided by the belief network model.

**d. GUI for displaying results and linking to specific summaries in PDQ will be built**

The graphical user interface has been remodeled and summarizes the data entered for a given patient, displays protocols in reverse order of appropriateness (i.e., most

appropriate trials are listed first), and suggests which variables should be asked next so that the maximum number of remaining trials can be triaged out of the list. The GUI is currently being tailored to lay users. We have presented the interface in local meetings and are in the process of changing the items that are not intelligible for the general public.

***Task 4.*** **Redesign of evaluation methods and interim analysis and system refinement, Months 12-24 (ACCOMPLISHED):**

**a. Evaluation methodology will be redesigned**

The evaluation strategy was redesigned to conform to the realities of the clinical services at Brigham and Women's Hospital and Dana Farber Cancer Institute. The need for unbiased oncologists to properly implement the proposed clinical trial is the critical point for its implementation in year 3. Retrospective data from Brigham and Women's Hospital has been obtained for preliminary testing of the model, with filing and approval from the Institutional Review Board.

**b. Interim analysis of the system using abstracted cases will be conducted**

These cases are being constructed based on actual retrospective data collected from the Brigham and Women's Hospital. We have collected data on 30 patients admitted to Brigham and Women's Hospital with a diagnosis of breast cancer stage IV. The items collected correspond to those on the WWW forms and were collected from paper charts.

**c. System will be refined in terms of belief network model and GUI given interim analysis results and internal user feedback.**

The first implementation of the belief network engine into the overall system has been finalized.

***Task 5.*** **Subject recruitment, abstraction of medical records, and creation of survey instruments for final analysis, Months 16-24 (partially ACCOMPLISHED):**

**a. Lay people ("patients") will be recruited**

We have contacted a number of individuals to help with the lay user interface, through contacts at the Harvard Medical School and the Massachusetts Department of Public Health Breast Cancer Program.

**b. Medical records will be abstracted and randomized**

Medical records have been abstracted by a research assistant. A simple computer-based interface was implemented to facilitate this task.

**c. On-line forms for recording selection of clinical trials for "patients" and providers will be built**

This forms are currently being implemented.

**d. Surveys for assessing "patient" and provider satisfaction with the system will be built**

We have added a human-computer interface specialist to the team to redesign the user interface and design and conduct the surveys.

**e. Primary care providers and oncologists will be scheduled for final experiments**

We are in the process of identifying collaborators for this task. We are targeting oncology fellows from the Dana Farber Cancer Institute. We have not anticipated the need for rewards, but this seems to be necessary to promote compliance. We are discussing this issue with our current collaborators.

The following tasks are anticipated for Months 24-36, and the plan has not been modified from the one stated in the Approved Statement of Work:

*Task 6.* **Evaluation experiments, Months 25-33:**

**a. Oncologists will assess system's performance**

**b. "Patients" will use the system and fill on-line forms and surveys**

**c. Primary care providers will use the system and fill on-line forms and surveys**

*Task 7.* **Final analysis and report writing, Months 34-36:**

**a. Final analyses of data from oncologists, "patients," and providers will be performed**

**c. A final report and manuscripts will be prepared**

In the next sections, we describe the design of FACTs, and illustrate with some screen samples from the existing prototype.

## *Design*

We have added a computer scientist with specialization in belief networks to our team (Dr. O. Ogunyemi). Dr. Ogunyemi has implemented belief networks for her PhD thesis in artificial intelligence at the University of Pennsylvania. She has taken the lead in all software engineering aspects of the system. Her first task was to improve the current code and enhance its documentation in order to facilitate team work. She has implemented all operators from Arden syntax into our expression evaluator software, and has adapted the code to incorporate all belief networks as defined in Netica proprietary files.

The summary below described the main components of the system. The main progress in this area was the inclusion of all operators from Arden syntax, the inclusion of classes for the belief network objects, and the updated documentation. The rationale for this design was given in the last progress report.

FACTS utilizes an evaluation engine called EV (Expression evaluator) to interpret Arden statements and expressions, including logical and temporal criteria. EV uses a lexer generated with flex 2.5.4 and a parser generated with Bison 1.25. Information about the clinical trial protocols, including the encoded criteria, is stored in XML documents. A separate XML parser is used to obtain the portion of an XML document containing the criteria encoded in Arden. The EV parser constructs an abstract syntax tree representing Arden statements and expressions. The evaluation of the abstract syntax tree follows an interpreter design pattern to recursively request the objects representing the nodes of the tree to interpret themselves and yield the result of the evaluation. The UML class diagram illustrating the object-oriented structure of EV was shown in the last progress report. Statements and expressions are related by inheritance to allow their participation in an interpreter or visitor design pattern. The object-oriented structure of FACTS was illustrated by a UML diagram in the last progress report. The information in the encoded criteria is maintained by the criteria store object. Arden variables may be evaluated upon demand with the variable evaluator object. Identifier evaluators, function evaluators, and type converters may be registered with the EV evaluator object, which it will consider using in the course of evaluating a statement or expression. With regard to evaluating functions, EV provides a base class called EVC_FunctionEvaluator that may be derived from in other projects. These may be registered with EV to be potentially used in evaluation. In the EVC_Evaluator::EvaluateFunction method, the "Evaluate" methods of evaluators pointed to by elements in fFunctionEvaluatorSeq are invoked, starting with the last EVC_FunctionEvaluator that was registered and working backward, until an evaluator is found that does not yield an unknown error or the first evaluator in the sequence is reached. If a suitable evaluator is found, its return value is returned. If no suitable evaluator is found, this method yields a lookup error. With regard to obtaining values of identifiers, EV provides a base class called EVC_IdentifierEvaluator that may be derived from in other projects. These may be registered with EV to be used in evaluation. In the EVC_Evaluator::GetIdentifierValue method, if an identifier is not known in the immediate context, the identifier evaluators in fIdentifierEvaluatorSeq are searched in reverse order, beginning with the last one to be registered. A particular evaluator is asked to determine the value of the identifier by calling the Evaluate method. If no error is generated, the identifier is considered to have been found. If there was an unknown error or lookup error, then searching continues. If there was another type of error, the routine fails. If after these lookup attempts the identifier has still not been found, the routine signals and lookup error. With regard to evaluating sentences in Arden in the form of an abstract syntax tree, EV mostly uses the interpreter design pattern. Work has been done on extending the capabilities of EV to use alternative evaluators for "where" expressions. The visitor design pattern was implemented to accomplish this.

Examples of the graphical user interface for FACTS were given in the last progress report.

# Key Research Accomplishments, Year 2

- Updated variables present in eligibility criteria for 85 protocols in PDQ
- Identified changes in protocol status
- Refined structure for representing and storing eligibility criteria and protocols
- Implemented syntax for representing eligibility criteria, allowing all operators from Arden syntax
- Improved parser for Arden syntax
- Refined graphical user interface to acquire patient data, summarize entries and display appropriate protocols
- Improved deterministic engine to match patient values against eligibility criteria
- Informally evaluated algorithm to rank protocols in reverse order of appropriateness for a particular case
- Redesigned evaluation as a clinical trial
- Collected and abstracted real cases from Brigham and Women's Hospital
- Started recruitment for clinical trial

# Reportable Outcomes

### *Manuscripts*

Ohno-Machado L, Boxwala AA, Wang SJ, Mar P. Decision Support for Clinical Trial Eligibility Determination in Breast Cancer. *Journal of the American Medical Informatics Association* 1999; Suppl 6: 340-4. (best paper award finalist)

Lacson R, Ohno-Machado L. A Comparative Trial of FACTS versus Usual Clinical Practice for Triaging Breast Cancer Patients. Technical Report, Decision Systems Group, Brigham and Women's Hospital and Harvard Medical School, 2000.

### *Abstracts*

Ohno-Machado L, Ogunyemi O, Kogan S. Decision Support for Clinical Trial Eligibility Determination in Breast Cancer. Abstract presented at the 2000 Breast Cancer Research Symposium of the Massachusetts Department of Public Health.

Wang SJ, Ohno-Machado L, Mar P, Boxwala AA, Greenes RA. Enhancing Arden syntax for clinical trial eligibility criteria. Proc *1999 AMIA Annual Fall Symposium,* Washington DC, 1999. Philadelphia: Hanley & Belfus. JAMIA (suppl) 1999: 1188.

Ohno-Machado L, Ogunyemi O, Le H, Greenberg S, Greenes RA. FACTS: Finding Appropriate Clinical Trials. The Internet and the Public's Health: Impact on Individuals, Communities and the World. Harvard School of Public Health and Harvard Medical School, May 30-31, 2000.

### *Presentations*

Poster presentation at the 2000 Breast Cancer Research Symposium of the Massachusetts Department of Public Health, 4/2000.

### *Informatics such as databases*

Database of Encoded Protocols available at
http://telmato.bwh.harvard.edu:8000/FACTS/data/

# Conclusions

We have accomplished the overall tasks of year 2 towards the construction of an automated system to automate patient eligibility match to suggest appropriate protocols for a specific patients. We have refined the prototype built in year 1. We have implemented an engine that deals with uncertain items and imputes appropriate values. We have designed a trial to formally evaluate the system and have collected real patient data from paper charts at Brigham and Women's Hospital. We have started to recruit subjects to our evaluation trial.

Our next steps are to (1) update the protocols to reflect status of trials and new additions, (2) modify the engine to account for associations among criteria families and (3) initiate formal evaluation of the system.

# References

Ohno-Machado L, Boxwala AA, Wang SJ, Mar P. Decision Support for Clinical Trial Eligibility Determination in Breast Cancer. *Journal of the American Medical Informatics Association* 1999; Suppl 6: 340-4.

Peleg M, Boxwala A, Ogunyemi O, Zeng Q, Tu S, Lacson R, Bernstam E, Ash N, Mork P, Ohno-Machado L, Shortliffe E, Greenes R. GLIF3. Submitted to the 2000 American Medical Informatics Association Fall Meeting. [Appendix 1].

Wang SJ, Ohno-Machado L, Boxwala A, Greenes RA. Enhancing Arden Syntax for Clinical Trial Eligibility Criteria. .Technical Report TR-199-02, Decision Systems Group, Brigham and Women's Hospital and Harvard Medical School, 1999.

Lacson R, Ohno-Machado L. A Comparative Trial of FACTS versus Usual Clinical Practice for Triaging Breast Cancer Patients. Technical Report, Decision Systems Group, Brigham and Women's Hospital and Harvard Medical School, 2000. [Appendix 2]

Appendix 1

# GLIF3

Mor Peleg, Ph.D.[1], Aziz A. Boxwala, M.B.B.S., Ph.D.[2], Omolola Ogunyemi, Ph.D. [2], Qing Zeng, Ph.D. [2], Samson Tu, M.S. [1], Ronilda Lacson, M.D. [2], Elmer Bernstam, M.D., M.S.E. [1], Nachman Ash, M.D.[2], Peter Mork, B.Sc. [1], Lucila Ohno-Machado, MD, Ph.D.[2], Edward H. Shortliffe, M.D., Ph.D.[1, 3], Robert A. Greenes, M.D., Ph.D.[2]

[1]Stanford Medical Informatics, Stanford University School of Medicine, Stanford
[2]Decision Systems Group, Harvard Medical School, Brigham & Women's Hospital, Boston
[3]Department of Medical Informatics, Columbia University, New York

The Guideline Interchange Format (GLIF) is a language for structured representation of guidelines that was developed for the purpose of facilitating sharing of clinical guidelines. GLIF version 2, which was published in 1998, enabled the modeling of a clinical guideline as a flowchart of structured guideline steps, representing clinical actions and decisions. However, the attributes of structured constructs were defined as text strings that cold not be parsed. Such guidelines could therefore not be used for computer-based execution. The next version of GLIF (known as GLIF3 is being designed to support computer-based execution. GLIF3 introduces several new constructs and structures existing GLIF2 constructs further to allow a more formal definition of decision criteria, action specifications and patient data. The new GLIF3 constructs enable guideline specification at three levels: a conceptual GLIF flowchart, a computable/parsable specification that can be validated for logical consistency and completeness, and an implementable specification that is appropriate for incorporation into particular institutional information system environments.

## Introduction

Clinical guidelines have been proposed as a way to standardize care in order to improve its quality and cost effectiveness. Structured, computer-based guidelines could be easily and conveniently delivered to the point of care in a way that enables decision support[1, 2]. Such guidelines could also provide workflow management support, quality assurance evaluation, and simulation of guideline execution for educational purposes[3].

Several approaches were purposed for creating computable guidelines. Lobach et al. describe a database schema based on a relational model for computerizing clinical practice guidelines using a hybrid of structured and procedural knowledge representation schemes[4]. The PROforma model is designed to support guideline dissemination in the form of expert systems, which assist patient care through active decision support and workflow management[5]. The Asbru language[6] can be used to create a guideline representation that includes the explicit intentions of the guideline's authors. Asbru can be used to represent complex, time-oriented actions and world states, as well as multiple intentions. The EON guideline model enables a

specification of a guideline trough a combination of modeling primitives, such as different types of decision-making mechanisms, flow of control constructs, actions and activities, and a distinction between the normal case and its exceptions[7]. The Arden syntax[8] is a language for creating and sharing medical knowledge in the form of independent modules, called medical logic modules (MLMs), each containing sufficient logic to make a single medical decision.

The task of creating clinical guidelines in the above computer-based form takes significant effort. Thus, it would be desirable to share clinical guidelines across institutions. Several logistical difficulties exist in sharing guidelines including differences in representation formats for guidelines and differences in computing platforms.

GLIF is a specification for structured representation of guidelines that developed by the InterMed Collaboratory in order to facilitate sharing of clinical guidelines[9]. The goal is not to allow interchange from one guideline formalism to another but to have guidelines authored and viewed by different software tools, shared in GLIF format and adapted for a variety of local uses[1].

The objective of the GLIF specification to provide a representation for guidelines that have the following characteristics: (a) Precise; (b) Non-ambiguous; (c) Human-readable; (d) Computable (in the sense that guidelines specified in GLIF may be used for computer-based decision support), and (e) Independent of computing platforms (thus enabling sharing of guidelines).

[1] In this sense, the word "interchange" in the expansion of the GLIF acronym (GuideLine Interchange Language) is a misnomer.

**Background**

Version 2.0 of GLIF (GLIF2) was published in 1998[9]. It consisted of the GLIF object model and the GLIF syntax. The GLIF model specified a guideline as a flowchart of temporally ordered clinical decision and action steps. Concurrency is modeled using branch and synchronization steps. The GLIF syntax, which was ODIF-based, specified the format of text files containing GLIF-encoded guidelines. This text file was used for sharing and interchange.

GLIF2 has been the basis for several implementations of guideline-based applications, including one in the Brigham's BICS information system[10]. Web-based applications for driving clinical consultations[3], and applications that search for eligible clinical protocols[11].

However, GLIF2 has several deficiencies that limit its widespread usability:

While important attributes of guideline steps have been defined, it lacks further structuring of these attributes. Thus, values of most attributes are specified as text strings. Thus it is not possible to use these guidelines for computer-based decision support.

GLIF2 lacks features to manage complexity in guideline flowcharts. Even relatively simple text guidelines, when specified in GLIF2 often end up as a large and complex flowchart. Such guidelines are more difficult for the user to edit or browse.

Guidelines in GLIF2 are difficult to integrate with heterogeneous clinical systems as it lacks features for mapping patient data references to entries in the electronic medical record.

The decision model used is limited and inflexible. Decisions are specified in a

Conditional Step that models decisions as an extended Boolean concept.

Only a limited set of low-level concepts are provided in GLIF2. Several important concepts such as for describing iteration, patient state, exceptional conditions, and events are lacking.

This paper presents GLIF3, a an evolving revision of GLIF, that attempts to overcome several of these limitations in GLIF2.

**Overview of GLIF3**

GLIF3 introduces several substantive changes to the object model and the syntax. In addition, GLIF3 enables guideline specification at three levels: a conceptual GLIF flowchart, a computable/parsable specification and an implementable specification.

Changes in the object model

The changes being made to object model included the definition of new constructs and further structuring of GLIF2 constructs.

Representation in UML

The GLIF3 model is described in the Unified Modeling Language (UML) class diagrams (12). Additional constraints on represented concepts are being specified in the Object Constraint Language (OCL), a part of the UML standard (12).

The GLIF2 specification was published in Interface Definition Language (IDL). A high-level object model was also defined.

Support for managing complexity in guideline instances

GLIF3 more fully defines a mechanism for specifying guideline steps iteratively through the nesting of subguidelines in action and decision steps. Figures 1 and 2 demonstrate how nesting is used to specify the details of the treatment action step of the Stable Angina guideline[13]. Since nesting allows grouping of parts of a guideline into modularunits (subguidelines), this is a mechanism that allows reuse of part of a guideline. Furthermore, the modularity of the guideline permits adaptation of a guideline to a specific institution by replacement or elaboration of well-defined sections of the guideline (i.e., replacing an action specified at a high-level with a detailed procedure).

GLIF3 also includes a capability for application-specific extensions in a scalable manner. These extensions can be built using the Macro_Step class. A macro can be used to represent application-specific constructs in a concise manner.

A capability to provide multiple views of the same guideline is also being added. This capability is provided through the use of filters that would collapse segments of the guideline from a default view of the guideline.

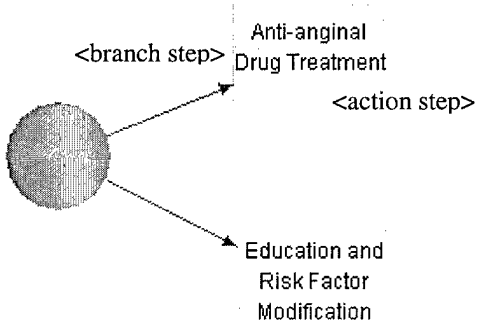



*Figure 1. The details of the Treatment algorithm, which is part of the Stable Angina Guideline*

Expression specification

A structured grammar for specifying expressions and criteria is added to GLIF3.

The grammar can be used for specifying logical criteria, numerical expressions, temporal expressions, and text string operations. The grammar is a superset of the Arden Syntax logic grammar[14]. It adds new operators such as "is a", "overlaps", "xor", "from now", "is unknown" and "at least *k* of ...".

Domain ontology support

In order to facilitate use of standard medical vocabularies and integration of shared guidelines into clinical information systems environment, GLIF3 uses a layered approach for referencing clinical terms. The core GLIF layer defines how medical data and concept may be represented and referenced by GLIF. The Reference Information Model (RIM) layer provides a semantic hierarchy for medical concepts that represent different classes of medical data. The hierarchy allows specification of the attributes of each class of medical data in an object-oriented fashion. Different RIM models (such as the HL7 RIM) may be used in different guidelines.. The medical ontology layer contains a term dictionary (e.g., UMLS) and can provide access to medical knowledge bases.

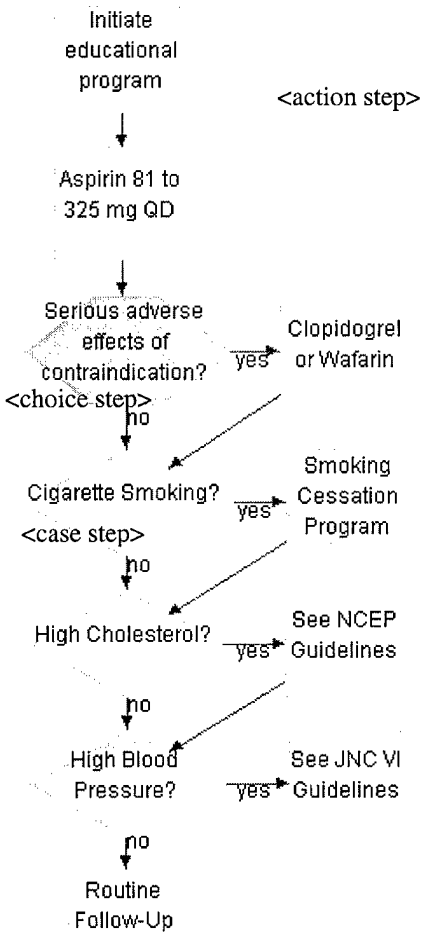



*Figure 2. The elaboration of the Education and Risk Factor Modification action step, shown in Figure 1. Case Step and Choice Step are described on the next page.*

In GLIF2, definitions of patient data items were provided through a Patient Data class contained in an Action Specification. The Patient Data class is now obsolete.

Flexible decision model

GLIF3 provides a flexible decision model through a hierarchy of decision step classes. The hierarchy can be extended further in the future to support different decision models. The decision hierarchy in GLIF3 distinguishes between decision steps that can be automated (case step) and ones that have to be made by a physician or other health worker

and cannot be automated (choice step). The decision hierarchy can be extended to model decisions that consider uncertainty or that consider patient preferences.

Extended action specification model

The action specification model has been extended to include two different types of actions: guideline-flow relevant actions (such as calling a subguideline, computing values for data) and clinically relevant actions (such as the recommendations). Clinically relevant actions reference the domain ontology for representations of clinical concepts such as prescription, laboratory test order, referral, etc.

Other new concepts

Representation for several new concepts were added to GLIF3. These include specifications for

Describing *Iterations* and conditions that control the iteration flow

Describing *Events* and triggering of guideline steps by events.

Describing *Exceptions* in guideline flow and associated exception-handling mechanisms.

Representing *Patient State* as another kind of guideline step (a node in the flowchart), in addition to the existing action, decision, branch, and synchronization steps. A patient state step serves both as a point of entry into the guideline and for labeling purposes.

A *Keyword Didactic* for adding keywords to a variety of constructs in guidelines

Corrections to branch and synchronization step

The branch and synchronization step have been modified to remove redundancy in descriptions of parallel pathways in the guideline flowchart.

Changes in the GLIF syntax

XML -based syntax

The proprietary ODIF-based syntax[15] in GLIF2 is being replaced with an XML-based syntax (16). A schema for the syntax is being developed.

Guideline Abstraction Levels

GLIF3 enables modeling guidelines at three levels of abstraction[17]:

Level A: conceptual flowchart. Guidelines that are specified in GLIF at this level can be used for browsing and reading in guideline viewing programs. However, these guidelines cannot be used for computation for decision support.

Level B: computable level. Guidelines specified at Level B may be validated for logical consistency and completeness (not correctness). For example, the syntax of expressions, definitions of patient data items and clinical actions, and flow of the algorithm would be specified at this level.

Level C: implementable level. At this level, guidelines are appropriate for incorporation into particular institutional information system environments. Thus, these guidelines may have non-sharable concepts defined within them.

**Discussion**

GLIF3 is evolving very rapidly. In particular, more work needs to be done on the specification of the data model and domain ontology. We are currently in the process of specifying several clinical guidelines, in the

three abstraction levels, in order to evaluate GLIF3. The full GLIF3 specification will be published on the Internet soon after in order to solicit comments from the user community.

In future versions of GLIF, we will explore structured representations for (1) specifying goals of guideline steps (18), (2) probabilistic models for decision-making (19), and (3) incorporation of patient preferences in decision steps (20).

Software tools are being developed for authoring, validating, viewing, and distributing guidelines. The tools are being implemented in Java to provide portability and use over the Internet.

**Bibliography**


1. Schriger DL, Baraff LJ, Rogers WH, Cretin S. Implementation of clinical guidelines using a computer charting system. Effect on the initial care of health care workers exposed to body fluids. JAMA 1997;278(19):1585-1590.

2. Lobach DF, Hammond WE. Computerized decision support based on a clinical practice guideline improves compliance with care standards. Am J Med 1997;102(1):89-98.

3. Boxwala AA, Greenes RA, Deibel SR. Architecture for a multipurpose guideline execution engine. In: Proc AMIA Symp 1999; 1999; 1999. p. 701-705.

4. Lobach DF, Gadd CS, Hales JW. Structuring clinical practice guidelines in a relational database model for decision support on the Internet. In: Proc AMIA Annu Fall Symp 1997; 1997; 1997. p. 158-62.

5. Fox J, Rahmanzadeh A. Desseminating medical lnowledge: the PROforma approach. Artificial Intelligence in Medicine 1998;14:157-181.

6. Shahar Y, Miksch S, Johnson P. The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines. Artificial Intelligence in Medicine 1998;14:29-51.

7. Tu SW, Musen MA. A Flexible Approach to Guideline Modeling. Stanford, CA: Stanford University; 1999. Report No.: Technical Report SMI-1999-0789.

8. Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden Syntax. Comput Biomed Res 1994;27(4):291-324.

9. Ohno-Machado L, Gennari JH, Murphy S, Jain NL, Tu SW, Oliver DE, et al. The GuideLine Interchange Format: A Model for Representing Guidelines. Journal of the American Medical Informatics Association 1998;5(4):357-372.

10. Zielstorff RD, Teich JM, Paterno MD, Segal M, Kuperman GJ, Hiltz FL, et al. P-CAPE: a high-level tool for entering and processing clinical practice guidelines. Partners Computerized Algorithm and Editor. Proc Amia Symp 1998:478-82.

11. Ohno-Machado L, Wang SJ, Mar P, Boxwala AA. Decision support for clinical trial eligibility determination in breast cancer. Proc Amia Symp 1999.

12. Object Management Group. Object Management Group (OMG) Unified Modeling Language (UML) Specification. In:. version 1.3 ed: Object Management Group; 1999.

13. American College of Cardiology/American Hospital Association/American College of Physicians-American Society of Internal Medicine. Guidelines for the Management of Patients with chronic Stable Angina. J Am Col Cardiol 1999;33:2092-2197.

14. Clinical Decision Support & Arden Syntax Technical Committee of HL7, inventor Arden Syntax for Medical Logic Systems, version 2.0. Draft revision. USA. 1999 July 7, 1999.

15. Pattison-Gordon E. ODIF:

Object Data Interchange Format. Boston, MA: Decision Systems Group, Brigham and Women's Hospital; 1996. Report No.: DSG-96-04.

16. W3C. Extensible Markup Language (XML). In:; 2000.

17. Greenes RA. Representation of Clinical Practice Guidelines to Facilitate Sharing; 1999 November 12, 1999. Report No.: IM-2000-4.

18. Shahar Y, Miksch S, Johnson P. An Intention-Based Language for Representing Clinical Guidelines. In: James J. Cimino E, editor. AMIA Annual Fall Symposium; 1996; Washington, D.C: Hanley & Belfus; 1996. p. 592-596.

19. Aliferis CF, Miller RA. On the heuristic nature of medical decision-support systems. Methods of Information in Medicine 1995;34:5-14.

20. Barratt A, Irwig L, Glasziou P, Cumming RG, Raffle A, Hicks N, et al. User's guide to the medical literature: How to use guidelines and recommendations about screening. JAMA 1999;281(21):2029-34.

Appendix 2 -

# A Comparative Trial of FACTS versus Usual Clinical Practice for Triaging Breast Cancer Patients

**Ronilda C. Lacson, MD and Lucila Ohno-Machado, MD, PhD**

## Specific Aims

FACTS (Finding Appropriate Clinical TrialS) is a computerized decision support system designed to enable physicians to identify the most adequate trials for a given patient. In this investigation, we plan to show that FACTS is more accurate in identifying the appropriate clinical trials for 40 patients with stage 4 breast cancer compared to usual oncologists' practice. Appropriate trials include all phase 2 and 3 trials for stage 4 breast cancer in the National Cancer Institute's (NCI) Physician's Data Query (PDQ) database for which each patient fulfills all eligibility criteria and is not excluded by the exclusion criteria. The null hypotheses are as follows: (1) There is no significant difference between the percentage of appropriate clinical trials identified by FACTS versus oncologists; and (2) There is no significant difference between the percentage of inappropriate trials identified by FACTS versus oncologists. Secondary analyses would focus on checking whether any significant differences detected above will be affected by limiting appropriate trials to only (1) those that are available in Massachusetts and (2) those that are available within the institution and it's affiliate hospitals.

## Background, Significance and Preliminary Studies

Historically, only 3-10% of eligible cancer patients are enrolled in clinical trials.[1, 2] Some of the reasons for this include physician factors such as lack of knowledge about what trials are being conducted and patient factors such as prejudice against scientific experiments, among other things.[1, 3] Low accrual in oncology trials is of particular concern because it can result in selective sampling that may compromise generalizability of the studies as well as increase the time needed to achieve adequate recruitment. This translates into increased costs, delayed evaluation and increased chance of trial arms becoming outdated.[4]

FACTS is a computer-based interactive system we built at the Decision Systems Group, Brigham and Women's Hospital (BWH) for accessing and evaluating breast cancer clinical trial

information. The system allows entry of details for a specific case using the Internet. It then indicates which trials the patient is eligible for. The rationale for selecting breast cancer trials was that this is the oncology domain containing the largest number of trials. We chose advanced stage breast cancer because we believe these patients might be more interested in participating in clinical trials. We decided to limit the set to phase 2 and 3 trials since these studies are further developed and are generally more available not only in academic centers. We found a total of 86 trials for which we have encoded 2188 eligibility and exclusion criteria.

Clinical decision support systems are software designed to directly aid in clinical decision making bout individual patients.[5] These have been successful in improving clinical performance and increasing access to guidelines, protocols and publications. [6-9] A few computer systems have been previously designed to assist in determination of clinical trial eligibility.[10, 11] Ohno-Machado et. al. previously developed a system for decreasing physicians' uncertainty regarding patients' eligibility for HIV treatment protocols.[10] There has not been any prospective evaluation of this system for accuracy and compared to usual physicians' practice. In the proposed study, we will assess the efficacy of FACTS in predicting appropriate trials for stage 4 breast cancer patients.

**Research Design and Methods**

Overall Strategy

We plan to assess the accuracy of FACTS and compare it to oncologists in predicting clinical trials for patients with stage 4 breast cancer. Clinical vignettes will be abstracted by the investigator from 40 randomly chosen patients with stage 4 breast cancer who were admitted to the BWH in 1995. For each patient, the list of appropriate trials will be collected from both a group of physicians who will be using FACTS and another group who have no access to FACTS. The primary analysis will be designed to compare the 2 groups for the cohort identified.

Study Sites

1) Brigham and Women's Hospital (BWH) – The BWH is a private teaching hospital affiliated with Harvard University offering primary, secondary and tertiary care.
2) Dana-Farber Cancer Institute (DFCI) – Dana-Farber/Partners Cancer Care is the collaboration in adult oncology among three world-renowned Harvard Medical School affiliated institutions - Dana-Farber Cancer Institute, Brigham and Women's Hospital and Massachusetts General Hospital. At the Longwood campus, inpatient care is provided at BWH while DFCI specializes in outpatient care. Patients are treated in 12

specialized centers; each devoted to a different kind of cancer. They have a large number of clinical studies available to patients and have a special support group particularly for metastatic breast cancer patients.

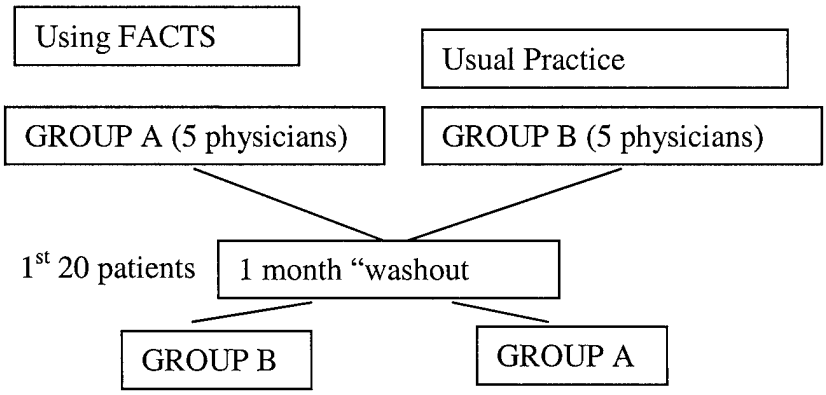
**Data Abstraction**

40 patients will be chosen at random from stage 4 breast cancer patients admitted to BWH in 1995. Permission will be obtained from the Institutional Review Board. The investigator will create clinical vignettes for each of these 40 patients by reviewing their medical records and data from the laboratory system. Only data needed to complete the FACTS form for predicting appropriate trials will be collected. (Appendix A) Data that are required for the form but not readily available will be left blank. Patient identifiers will not be included in the vignettes. Subsequently, a list of all appropriate clinical protocols for each patient will be created based on the patient's meeting all eligibility criteria and not being excluded by the exclusion criteria by each of the 86 protocols. This individualized list will serve as the "gold standard". For the secondary goal of the study, 2 more sets of gold standard lists will be generated for each patient, taking into account only those trials that are available within Massachusetts and those that are available within DFCI and its affiliates.

**Study Design**

10 oncologists from the Dana-Farber Cancer Institute will be randomly assigned to 2 groups using a computer randomization procedure. The first group will use FACTS to enumerate all trials for which each of the first 20 patients identified is eligible to participate in. The second group will just be allowed to name all trials that they think is appropriate for each of the same 20 patients given to the first group using whatever methods they use in practice (Internet, Medline, local network information, etc.). The patient cases will be arranged randomly to minimize easy comparisons for answers between participants. We will attempt to balance these physicians by classifying them by rank (attendings vs. fellows) before allocation and then randomizing within the subgroup. Assuming that one group performs comparatively better than the other group, we want to make sure that rank is not a confounder. After they finish listing trials they deem are appropriate for each patient; all participants will be given one month when nobody can have access to FACTS. Subsequently, a second phase will be started where the 2 groups of physicians will be crossed-over and FACTS will be available to those who did not have it in the first phase. This will allow us to

compare individual physicians' performance with and without FACTS and address problems with having less computer competent physicians in one group. A second set of 20 randomly arranged patients would again be distributed to each physician. They will again be asked to identify appropriate trials for each patient. In this design, equal distribution of patients to each intervention group is assured since the same vignettes will be analyzed, assuring that apart from the availability of the system in the decision-aid group, everything else about the case is the same. The schema is depicted below:




The presence of unbalanced Hawthorne effect is probably not as big a problem in this case since we will not allow the physicians to modify the output of the system. Only errors in data entry and understanding of the clinical cases may affect the result of the system. The errors should be minimal since we have tested the system extensively for robustness. Novelty of the technology will be addressed by giving physicians ample time to familiarize themselves with the system and providing technical support prior to their work on the cases. Illicit use of the decision-aid in the non-FACTS users will be addressed by granting access to the system only for appropriate user names and passwords.

**Data Collection and Definition of Outcome**

The list of trials for each patient given by each oncologist will be collected. The resulting list for each patient will be compared to the "gold standard" as previously identified. A separate investigator who is blinded to the source of the list (from a FACTS vs. non-FACTS user) will compute 2 measures for each patient:

(1) Appropriateness Ratio (AR) = # of trials chosen that agree with the gold standard

$$\text{(2) Inappropriateness Ratio (IR)} = \frac{\text{\# of trials chosen that are not in the gold standard}}{\text{\# of all inappropriate trials for the vignette}}$$

Based on these, the average AR for the 5 oncologists who used FACTS to identify appropriate trials for each patient ($AR_{FACTS}$) can be identified **for each patient** as well as for the 5 physicians who did not use FACTS ($AR_{NO\text{-}FACTS}$). Similarly, we can do the same for the average IR's ($IR_{FACTS}$ and $IR_{NO\text{-}FACTS}$).

(1) $AR_{FACTS}$ = [Σ AR of 5 physicians for same patient (using FACTS)]/5

(2) $AR_{NO\text{-}FACTS}$ = [Σ AR of 5 physicians for same patient (not using FACTS)]/5

(3) $IR_{FACTS}$ = [Σ IR of 5 physicians for same patient (using FACTS)]/5

(4) $IR_{NO\text{-}FACTS}$ = [Σ IR of 5 physicians for same patient (not using FACTS)]/5

**Data Analysis**

A comparison of performance between oncologists who use FACTS vs. those who didn't on the exact same patients lend this study to a paired analysis. For each of the 40 patients in the study, the differences in $AR_{FACTS}$ and $AR_{NO\text{-}FACTS}$ will be calculated (dependent variable). Differences between values are usually normally distributed even if the parent distributions are not, thus leading to use of a paired t-test for analysis. The null hypothesis would be that the difference is 0 if the AR's are the same for FACTS vs. non-FACTS users (Hypothesis 1). Similarly, differences will be taken between $IR_{FACTS}$ and $IR_{NO\text{-}FACTS}$ and the same analysis will be performed (Hypothesis 2). Appendix B demonstrates the format of the expected results. Contamination of the results in the second phase ("carry-over effect") will be addressed by comparing the AR's and IR's between similar physicians while they were using FACTS and while they were not, in each of the phases. A 2-sample t-test comparing results for the 1st group of 20 patients vs. that for the 2nd group of 20 patients (for each of the 2 physician groups, Group A and Group B) will be performed. If found to be significant (meaning there is carry-over effect), we will disregard the second phase of the trial.

To address the secondary goals of the study, we will perform the same calculations above with the gold standard varied as described in the Data Abstraction section. The dependent variable (differences between AR's and IR's) for each patient can be compared to the values obtained above.

Since these variables are for the same patient, a paired t-test can again be used to check the influence of limiting the gold standard to local trials alone.

Statistical Power

We performed power calculations for the one sample t-test. The effect should be greater when AR's are compared as opposed to IR's. To be conservative, the power to detect a 0.10 difference from the null (with 0.20 standard deviation) with a sample size of 40 is between 0.85 and 0.95. If we have moderate effect of .20, then the power increases to above 0.995. There is more power since we are utilizing a paired analysis.

Work Plan

The data abstraction phase will be completed in 2 months, including the preparation of the gold standard. The randomization of the physicians to the use of FACTS will also be done in this period. The last week will be used to train those assigned to using FACTS. The oncologists will be given 2 weeks to complete the first phase of the study. There will be a one-month washout period, the last week of which will be used to train the FACTS users for the second phase. The oncologists will again be given 2 weeks to complete the second phase of the study. Data analysis and development of reports will be done in 2 months for a total of 6 months for study completion.

Significance and Generalizability

In a regular field trial, a system's impact on decision-making depends on whether physician's act upon the system's recommendations. This addresses issues of how to modify physician practices and behavior. We are not addressing this presently and so it is appropriate that we have limited the study to case vignettes where physicians would hopefully be more exact and knowledge-based in their recommendations for appropriate trials. They will be less influenced by reimbursement, concerns about losing patients to other institutions and potential legal implications.[12] Workflow issues should also be addressed separately and will not be covered in this study. Other than the biases stated in the data collection and analysis portions of the study, it is important to consider that performance in an artificial setting is not the same as when one is in front of a patient. Oncologists from DFCI may also be more aware about clinical trials than most physicians. This will bias the result towards the null and finding an effect would be an underestimate of what one would really observe. Generalizability of the effect of FACTS to other domains (non-breast cancer patients) may

depend on other factors such as complexity of eligibility criteria, average available number of trials per disease (or stage), physicians' general knowledge and behavior about clinical trials and decision support systems, and patient preferences. We believe, however, that this tool can increase physicians' knowledge about available and appropriate clinical trials for their patients, thus increasing participation in trials. We also plan to make it available for patients to increase their participation in decisions regarding their health. This study can help inform physicians about the scope and performance of this system in identifying appropriate clinical trials.

**References**


1. Winn RJ. Obstacles to the accrual of patients to clinical trials in the community setting. *Seminars in Oncology*, 1994 Aug, 21 (4 Suppl 7): 112-7.
2. Schain WS. Barriers to clinical trials, Part II: Knowledge and attitudes of potential participants. *Cancer*, 1994 Nov, 74 (9 Suppl): 2666-71.
3. Mansour EG. Barriers to clinical trials, Part III: Knowledge and attitudes of health care providers. *Cancer*, 1994 Nov, 74 (9 Suppl): 2672-6.
4. Morrow GR, Hickok JT, Burish TG. Behavioral aspects of clinical trials. *Cancer*, 1994 Nov, 74 (9 Suppl): 2676-82.
5. Randolph AG, Haynes RB, Wyatt JC, Cook DJ, Guyatt GH. User's guide to the medical literature. *JAMA*, 1999 Jul, 282:67-74.
6. Sullivan F, Mitchell E. Has general practitioner computing made a difference to patient care? A systematic review of published reports. *BMJ*, 1995 Sep, 311: 848-52.
7. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med*, 1994 Jan, 120: 135-42.
8. Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *JAMIA*, 1996 Nov/Dec, 3: 399-409.
9. Overhage JM, Tierney WM, Zhou X, McDonald CJ. A randomized trial of "corollary orders" to prevent errors of omission. *JAMIA*, 1997 Sep/Oct, 4: 364-75.
10. Ohno-Machado L, Parra E, Henry SB, Tu SW, Musen MA. AIDS2: A decision-support tool for decreasing physicians' uncertainty regarding patients' eligibility for HIV treatment protocols. *SCAMC*, 1993, 429-33.